



Do not hesitate! – Unless you do it shortly or nasally: How the phonetics of filled pauses determine their subjective frequency and perceived speaker performance

Oliver Niebuhr¹, Kerstin Fischer²

¹Mads Clausen Institute, Centre for Industrial Electronics, University of Southern Denmark

²Dept. of Design and Communication, University of Southern Denmark

oniebuhr@mci.sdu.dk, kerstin@sdu.dk

Abstract

In this paper, we test whether the perception of filled-pause (FP) frequency and public-speaking performance are mediated by the phonetic characteristics of FPs. In particular, total duration, vowel-formant pattern (if present), and nasal-segment proportion of FPs were correlated with perceptual data of 29 German listeners who rated excerpts of business presentations given by 68 German-speaking managers. Results show strong inter-speaker differences in how and how often FPs are realized. Moreover, differences in FP duration and nasal proportion are significantly correlated with estimated (i.e. subjective) FP frequency and perceived speaker performance. The shorter and more nasal a speaker's FPs are, the more do listeners underestimate the speaker's actual FP frequency and the higher they rate the speaker's public-speaking performance. The results are discussed in terms of their implications for FP saliency and rhetorical training.

Index Terms: filled pause, rhetoric, hesitation, German.

1. Introduction

Filled pauses (FPs) such as "err", "uh", "um", and "mmh" [1], also referred to as hesitation markers, generally have a bad reputation. For instance, [2] finds that only 4% of her 105 informants assume that filled pauses can be helpful, and 73% answer that they try to avoid them. Often, speakers who use FPs are judged as unconcentrated, nervous, less knowledgeable [3] or even as lying [4]. These folk notions are in contrast to scientific studies.

Early psycholinguistic works on FPs set out to find direct connections between the occurrences of FPs and particular linguistic planning processes [8]. For instance, determining whether FPs occur more often at utterance onsets or within utterances was expected to provide evidence for sentence planning and word search activities respectively [9,10]. Such a direct window into speech processing could not be established. Instead, many pragmatic functions of FPs were identified, such as mitigating otherwise potentially impolite utterances [11,12] and displaying concern for the addressee. Other studies show that FPs cue important functions regarding the structuring of information and that these cues are reliably made use of by listeners. In this way, FPs can facilitate speech processing [5,6] and even content memorization [6,7].

For example, nasal FPs (such as "um") indicate a longer upcoming delay than non-nasal FPs (such as "uh" or "err"), see [1]. Connected to this systematic difference in FP realization and distribution is the communicative function that nasal FPs are used to mark (i.e. introduce) larger discourse units, cf. also [3]. Relatedly, [3] reports differences regarding the size of repaired units. With "um" as a repair marker, often

the whole constituent is repaired, whereas with "uh", generally only one word is repaired, i.e. corrected by the speaker. To sum up, empirical evidence suggests that FPs are actually better than their bad reputation.

More recent works concentrate on the distribution of FPs across speakers and speech situations. For instance, [13] notices considerable interpersonal differences between the number of FPs per minute by speaker, as well as differences depending on who the speaker is talking to and in what kind of context. Analyses of large corpora carried out by [14] and [15] find a significant increase in FP use among younger speakers. Furthermore, they find women to lead the process, especially with an increasing use of nasal variants like "um".

In summary, FPs are to a certain degree systematically used, but there are also large inter-individual differences in the frequency of use of FPs and in their phonetic characteristics. At the same time, both the frequency of occurrence and the phonetic characteristics of FPs are so stable within a speaker that FPs are used in forensics for speaker comparison and even for speaker identification [16,17,18].

Although FPs fulfill important functions and, thus, unlike stated in most rhetorical manuals [19,20], need not be banned from a speaker's utterances, too many or too prominent FPs can be disruptive [1]. This may mean that speakers can profit from either using fewer FPs or from realizing them such that they are less salient. In the area of phonetics, salience is measured in terms of perceptual prominence. The most important acoustic trigger for perceptual prominence, f_0 , is no major source of prominence in FPs, though. This is because FPs usually show hardly any dynamic f_0 patterns (i.e. rises/falls) that could trigger different degrees of prominence [13,16,21]. Instead, prominence differences between FPs are primarily caused by duration and intensity – with intensity being inherently controlled by two factors: vowel quality (front, open, un-rounded vowels have an inherently higher intensity, [22]) and the FP's nasal-segment proportion (nasals have an inherently lower intensity than vowels [22]; i.e. the higher this nasal proportion the lower is the overall intensity of the FP).

Against the rhetorical background of fluent speech production, the present paper focuses on the number of FPs. That, we exclude, for now, all functional and distributional aspects of FP realization. We ask how closely correlated the actual *physical number* of FPs in the speech signal is with the *subjective number* of FPs perceived by listeners.

Specifically, we examine whether listeners systematically *overestimate* the number of FPs the longer and less nasal the FPs are realized – and/or whether they systematically *underestimate* the number of FPs the shorter and more nasal the FPs are realized. In addition, we address the question whether shorter and more nasal FPs are better for a speaker's perceived rhetorical skills.

2. Method

2.1. Speakers

The speech data for the perception experiment came from 68 speakers, 36 females and 32 males. They were between 27 and 58 years old, average age 36.7 years. The speakers participated in "Persuasive Communication" courses that are regularly offered by the first author to professionals with managerial responsibilities or intensive customer contact. In these courses, short presentations of 2-5 minutes are practiced and recorded digitally. The speech data used here represent a subset of these presentation recordings. The subset is limited to the baseline presentations that are recorded course-initially prior to training. The presentations are given in German by managers with a medium public-speaking experience (<5 yrs).

2.2. Stimuli

A section of about one minute was extracted from the middle of each speaker's presentation, with attention to prosodic and syntactic completeness. These 68 one-minute excerpts served as stimuli. Presentation beginnings were omitted to exclude "warm-up" phenomena, endings were omitted to exclude any company names or slogans and frequently recited closing phrases, which typically lack FPs and are also in other respects not representative of the speakers' spontaneous speech.

2.3. Experiment and procedure

The 68 stimuli were integrated in a perception experiment using PRAAT MFC [23] with listener-individual stimulus randomization and no stimulus repetition. Two groups of listeners participated in this experiment in individual sessions. At the beginning of a session, listeners received verbal and written instructions on the basis of a prepared, constant text.

One group of listeners (group A) was informed that the experiment would be about FPs like "uh," "um," "err," and "mmh" that speakers use regularly in everyday communication. A set of FP examples were played to them, selected to cover the full range from vocalic over partly-nasal to fully-nasal long and short FPs of male and female speakers. None of these examples was included in the 68 stimuli. It was stressed that silent pauses would not count as FPs.

After the examples had been played, listeners in group A were instructed to listen to the one-minute recordings of all 68 different speakers and to estimate at the end of each recording how many FPs the speaker produced in that minute. Listeners were told not to listen analytically. It would not be their task to determine the correct FP number. Rather, the experimenters would be interested in the perceived order of magnitude, i.e. a subjective FP count, for which there is no right or wrong answer anyway. Still, to prevent the listeners from any conscious FP counting, they were given, prior to each stimulus, a shuffled set of playing cards – only number cards. Listeners were asked to sort these cards by suit and ascending numbers while listening to the played stimulus. Pilot tests showed that this extra task effectively suppressed any FP counting – silently or using fingers. When a stimulus ended, listeners entered the estimated FP count as an integer on a prepared sheet. Then, the set of number cards was exchanged for a new one, and listeners proceeded to the next stimulus by pressing a button, starting once again with their combined sorting-listening task and the subsequent FP frequency estimation.

The other group of listeners (group B) received the same 68 stimuli, but a different task. Group B's task was listen to

each stimulus in a content-oriented and holistic way, as if listening to a lecture, and then rate afterwards how skilled they perceived the speaker's presentation performance to be on a 10-point scale from 0='extremely bad' to 10='absolutely excellent'. Fleshing out the term 'presentation performance' further, listeners were told a 10 would mean that they continuously paid attention, enjoy listening, that it was not difficult to follow the speaker, and that they would want to continue listening to the speaker. As in group A, group-B listeners responded by entering integers on a prepared sheet of paper.

The listeners of groups A and B heard the stimuli via headphones at a constant, pre-set volume level. The experiment was conducted in a quiet room at Kiel University. For group-B listeners, the experiment took about 80 minutes. Group-A listeners needed 10 minutes longer (90 minutes) due to the playing-card instructions and exchanges.

2.4. Listeners

Since the experiment was long and carried out in individual sessions, only 29 listeners took part in total. Group A consisted of 13 listeners, 5 males and 8 females. Group B consisted of 16 listeners, 9 males and 7 females. All of the 29 listeners were engineering students. They were German native speakers, like the speakers in the stimuli, and between 23-29 years old (\bar{x} 25.6 yrs group A; \bar{x} 23.9 yrs group B). No listener suffered from a hearing disorder or had experience with public-speaking or persuasive communication training.

2.5. Analysis of FPs

Besides collecting listener estimations and ratings, the FPs contained in the 68 stimuli were also analyzed regarding four measures: First, the actual (objective) number of FPs was counted per stimulus/speaker, based on the acoustic signal. Second, the mean duration of a speaker's FPs was measured. Third, the mean vowel quality with which a speaker realized his/her FPs was determined. Following [24], vowel openness was measured in terms of the F1-f0 difference. Vowel quality along the front-back vowel dimension (including lip rounding) was measured in terms of the F2-F1 difference. All Hz values were converted to Bark. Fourth, it was determined through spectrogram analysis whether each individual FP of a speaker was realized primarily as a nasal or as a vowel segment. On this basis, it was calculated per speaker – in percent – how many of his/her FPs were realized primarily as nasal segments. For example, a %-Nasal value of 66.6 % would indicate that, on average, two out of three FPs within a speaker's one-minute presentation excerpt consisted mainly ("um") or entirely ("mmh") of a nasal segment.

A total of 430 FPs were analyzed across the 68 speakers. All acoustic analyses were carried out with PRAAT [25], as it was found to be better for formant measurements than alternative signal processing tools [26].

3. Results

3.1. Acoustic features of the 68 speakers' FPs

Figure 1 shows how the mean FP durations are distributed across the 68 speakers. The largest group of speakers used FPs with mean durations of 300-400 ms in their presentations. The second largest speaker group used very short FPs of 100-200 ms (or <100 ms). The third largest speaker group fell in between the other two groups with mean FP durations of 200-300 ms. Additionally, for a number of speakers FP durations

were on average longer than 400 ms. However, above 400 ms mean durations of FPs were also a lot more variable across speakers so that no further subgroups of speakers emerged. Therefore, all speakers with mean FP durations >400 ms together constituted a fourth group of speakers.

In summary, four groups of speakers were formed based on the speaker's individual FP durations: Group 1 with short FP durations (<200 ms, n=12), Group 2 with medium-short FP durations (200-300 ms, n=15), Group 3 with medium-long FP durations (300-400 ms, n=26), and Group 4 with long FP durations (>400 ms, n=15). Note that the overall range and distribution of FP durations match well with measurements of previous studies [16, 27, 28,29,30].

Based on the four FP-duration groups, a four-way MANOVA was conducted with the fixed factors F1-F0 (Bark), F2-F1 (Bark), Total FP Count (of a speaker), and %-Nasal (i.e. % of a speaker's FPs realized mainly as a nasal segment). Speaker was included as a covariate. FP duration (groups 1-4) had an effect neither on Total FP Count nor on F1-F0 and F2-F1, i.e. on the FPs' vowel quality. Figure 2 shows that, overall, the vowel qualities of FPs varied a lot between speakers within the – for German expected – triangle of [e]-like, [œ]-like, and [ə]-like sounds [18,31].

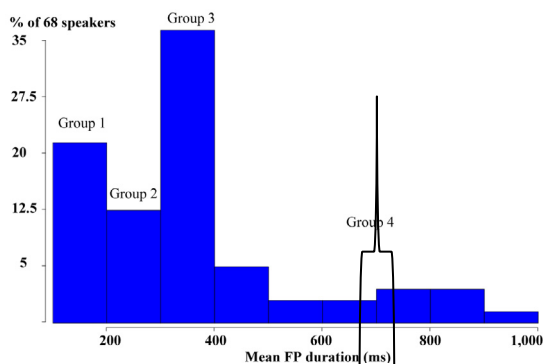


Figure 1: Mean FP durations across the 68 speakers and the resulting 4 speaker groups (N=68).

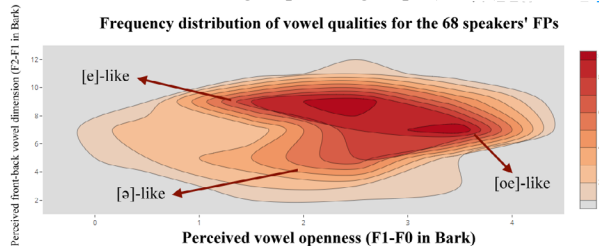


Figure 2: Density plot of mean F1-F0 and mean F2-F1 values of each of the 68 speakers' FPs (N=68).

FP duration had an effect on %-Nasal ($F[3,64]=3.199$, $p=0.044$). FPs with durations > 400 ms contained more or longer nasal segments than FPs with durations < 200 ms. This agrees with the previous finding that FPs with nasal portions are longer than purely vocalic FPs [1,29]. However, note with respect to 3.2 below that the difference underlying the effect of %-Nasal is relatively small (20.2 % between groups 1 and 4), much smaller than the overall between-speaker variation (10-100 %).

The effect of the covariate Speaker was also significant. Data inspections showed that one reason for this significance (besides a generally tremendous between-speaker variation) is speaker sex. Male speakers produced more FPs than female speakers, and female speakers produced more nasal FPs than

male speakers. Both is consistent with results from previous studies, e.g., [13,15,30]. Unlike [21], we found no effect of speaker sex on the vowel quality in FPs, but this is probably because we measured vowel qualities as speaker-normalized spectral distances and in terms of perception-based bark values rather than in terms of absolute Hz values.

The Total FP Count varied among the 68 speakers between 1 and 13 FPs per minute, which is within the range found in previous works. The average Total FP Count was 6.33 FPs per minute, which is typical for German speakers [16]. The value (6.33) is even identical to that reported in the German FPs analysis of [27].

3.2. Perceptually estimated FP count and speaker performance ratings

In order to determine the relationship between perceptual evaluation and the acoustic-phonetic properties of FPs, a two-way MANOVA was conducted, again with the four FP-duration groups as independent variable. Fixed factors were the difference between the estimated and the actual FP count per speaker and the speaker performance rating. Speaker was added as a covariate.

We found an effect of FP duration on both the estimated-to-actual FP count difference ($F[3,64]=15.483$, $p<0.001$) and the speaker-performance rating ($F[3,64]=5.375$, $p=0.002$). The covariate Speaker was significant, too.

Closer inspections of the significant effects revealed a significant correlation between FP durations and the estimated-to-actual FP count difference ($r[66]=0.611$, $p<0.001$), see Figure 3. For speakers with short FPs (< 200 ms), listeners underestimated the actual number of FPs in their speech (negative values on the y axis in Fig.3). The opposite applies to speakers with long FPs (> 400 ms). Listeners overestimated the actual number of FPs in their speech. The most accurate estimations of a speaker's actual FP counts were made for those speakers whose FPs were on average about 250-350 ms long – which is close to the overall mean FP duration in the present and preceding studies [13,28,29].

Furthermore, Figure 4 shows that a speaker's performance rating is significantly negatively correlated with his/her mean FP duration ($r[66]=-0.313$, $p=0.009$). The longer the FPs of a speaker are the less competent s/he sounds in the ears of listeners. An increase in FP duration by 400 ms lowers a speaker's perceived performance by about 1 scale point; or, in other words, compared to those speakers with the shortest FPs, those speakers with the longest FPs sounded 50 % less competent. Also note the dependency of the two listener variables: The estimated number of FPs was strongly negatively correlated with speaker performance ($r[66]=-0.417$, $p<0.001$). The fewer FPs listeners perceived in a speaker's speech, the higher they rated the speaker's presentation performance.

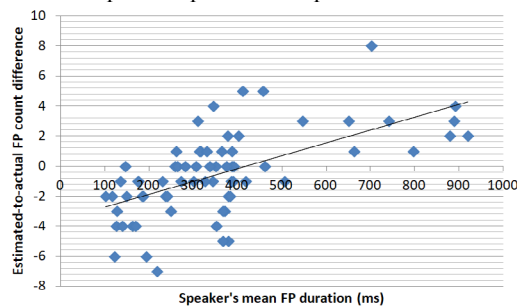


Figure 3: Over/underestimation of FP count depending on the speaker's mean FP duration.

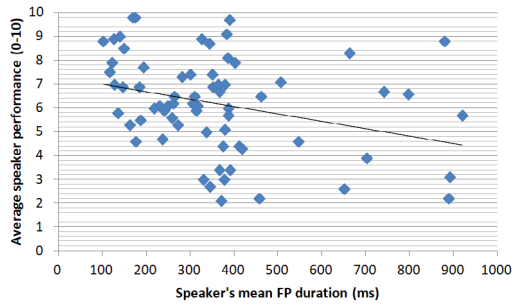


Figure 4: *Speaker performance rating depending on the speaker's mean FP duration.*

Based on these clear results, we extended the correlation analysis to the %-Nasal variable and found a similarly clear but inverted pattern, see Figures 5-6. The more a speaker's FPs were dominated by a nasal segment, the more strongly did the listeners underestimate the actual number of FPs in the speaker's speech ($r[66] = -0.391$, $p = 0.001$); and the more a speaker's FPs were dominated by a nasal segment, the higher was the speaker's performance rating ($r[66] = 0.461$, $p < 0.001$).

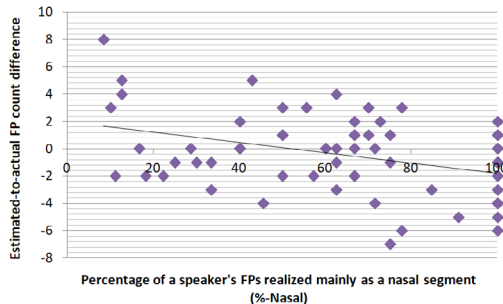


Figure 5: *Over/underestimation of FP count depending on the speaker's % of mainly nasal FPs.*

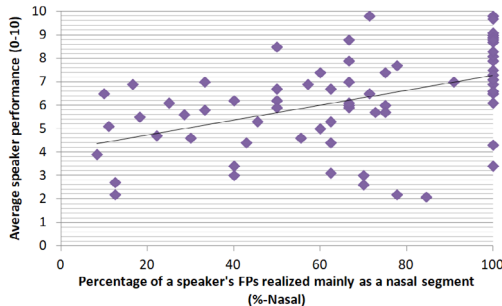


Figure 6: *Speaker performance rating depending on the speaker's % of mainly nasal FPs.*

4. Discussion and Conclusions

The results of the acoustic analysis and the perception experiment suggest that listeners are generally not good at estimating the number of FPs produced by speakers. Listeners mostly over- or underestimate how many FPs speakers produce, and – this is the decisive point – the direction as well as the extent of this misestimation are in speaker's own hands.

It depends on the average individual characteristics of FPs if and to what extent their frequency is over- or underestimated. We acoustically analyzed the total duration of FPs, their vowel quality, and the degree to which they were realized as a nasal sound segment. In accordance with previous studies, we found strong inter-individual variation along all three acoustic dimensions, while, at the same time, speaker-specific FP profiles emerged as well. However, only two acoustic dimensions

were found to be correlated with the misestimated number of FPs: the FPs' total duration and the degree to which they were realized as a nasal segment. Vowel quality proved irrelevant for our research questions, perhaps because the majority of speakers realized FPs mainly as a nasal so that the vowel portion was too short for a separate significant effect on the estimated number of FPs. A follow-up study should investigate this possibility, ideally based on a language whose speakers have a general preference for using vocalic FPs [27].

Regarding the two acoustic dimensions that proved relevant here, our data show that listeners are best at estimating the actual number of FPs when they literally show an "average" realization, i.e. a duration of between 300-400 ms and a nasal proportion of about 50%. We also see in our data that the listeners' FP estimation can, for a one-minute speech excerpt, deviate by 6-8 FPs from the actual number of FPs upwards and downwards. This is in the same order of magnitude as the FPs' average frequency of occurrence per minute. Thus, listeners' misestimations can be in the range of 100% of the number of FPs actually produced by a speaker.

At this point, our results become important for a rhetorical perspective on FPs since they show that the same phonetic FP features that make listeners underestimate the actual number of FPs also make them rate a speaker better in terms of his/her speech performance (public-speaking skills). Thus, our clear rhetorical recommendation is not to try to generally avoid FPs while giving a speech (also for reasons given in [1,3]), but to try to produce them mainly as a nasal and with a total duration of at most a single syllable (400 ms, preferably shorter). Note that we assume here that the two correlations are linked by perceptual prominence, i.e. FPs with a short duration and a largely nasal realization are perceived as less prominent and hence more often go unnoticed by listeners.

Although this is a plausible and obvious explanation, its individual assumptions need to be examined in follow-up studies, since there are other possible accounts for the two correlations detected. For example, nasal proportion might not have had a positive effect on the perceived speaker performance because it lowered the FPs' perceptual prominence but because nasal FPs simply sounded more pleasant in the ears of listeners than vocalic FPs. Various studies revealed a general sociolinguistic trend [14,15] that nasal-dominated FPs are "in" and that their use is generally increasing across languages, including German. In addition, according to a study by [32], FPs may be influenced by the speakers' personalities, which, in turn, could then be reflected in ratings like our measure of speaker performance. What also argues against a mere prominence-related effect of the nasal proportion in FPs is that nasal FPs occur more often with longer pauses than vocalic FPs [1]. These longer pauses should make nasal FPs more rather than less prominent.

To sum up, this discussion shows that, on the one hand, there are still many questions about the origins of the effects found here and their generalization about other languages and speaking situations. On the other hand, the found effects suggest that a separate focus on FPs in rhetorical training can be useful and that phonetic tools for signal processing and visualization may support effective rhetorical training.

5. Acknowledgements

The work reported on was partially funded by the Danish Council for Independent Research under Grant No. 4180-00359A. Thanks are also due to Nathalie Schümchen and Jan Michalsky for many inspiring discussions on FPs in speech.

6. References

- [1] Clark, Herbert H. & Fox Tree, Jean E. (2002): Using *uh* and *um* in Spontaneous Speaking. *Cognition* 84: 73-111.
- [2] Fox Tree, Jean E. (2002): Interpreting Pauses and *Ums* at Turn Exchanges. *Discourse Processes* 34,1: 37-55.
- [3] Fischer, K. 2000. From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Markers. Mouton de Gruyter.
- [4] Smith, Vicki L. and Herbert H. Clark. 1993 On the Course of Answering Questions. *Journal of Memory and Language* 32: 25–38.
- [5] Fox Tree, Jean E. (2001): Listeners' uses of *um* and *uh* in speech comprehension. *Memory and Cognition* 29, 320-326.
- [6] Corley, M. & Hartsuiker, Robert J. (2003). Hesitation in speech can ...*um*... help a listener understand. Proceedings of the 25th Annual Meeting of the Cognitive Science Society.
- [7] Fruehwald, Josef (2016): Filled Pauses as a Sociolinguistic Variable. U. Penn Working Papers in Linguistics 22.2, p. 41-49.
- [8] Hawkins, P.R. (1971): The Syntactic Location of Hesitation Pauses. *Language and Speech* 14: 277-288.
- [9] Goldmann-Eisler, Frieda (1958): Speech Production and the Predictability of Words in Context. *Quarterly Journal of Experimental Psychology* 10, 96.
- [10] Howell, Richard W. & Vetter, Harold J. (1969): Hesitation in the Production of Speech. *The Journal of General Psychology* 81: 261-276.
- [11] Levinson, Stephen (1983): *Pragmatics*. Cambridge: Cambridge University Press.
- [12] Schegloff, Emanuel A. (2010): Some Other “Uh(m)”s. *Discourse Processes*, 47(2), 130-174.
- [13] Shriberg, Elizabeth (2001): To ‘errr’ is human: Ecology and Acoustics of Speech Disfluencies. *Journal of the International Phonetics Association* 31, 1: 153-169.
- [14] Fruehwald, Josef (2016): Filled Pauses as a Sociolinguistic Variable. U. Penn Working Papers in Linguistics 22.2, p. 41-49.
- [15] Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J. and Liberman, M. (2016). Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change*, 6(2), 199-234.
- [16] Braun, A. & A. Rosin (2015). On the speaker-specificity of hesitation markers *Proc. 18th International Congress of Phonetic Sciences, Glasgow, UK*, 731-734.
- [17] King, J., Foulkes, P., French, P., & Hughes, V. (2013). Hesitation markers as parameters for forensic speaker comparison. *Proc. IAFPA conference, Tampa, Florida, USA*.
- [18] Künzel, H.J. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics* 4, 48–83.
- [19] Ni, P. (2017). How to Reduce “Um’s” and “Ah’s” (fillers) when Speaking. *Psychology Today* 2017, URL: <https://www.psychologytoday.com/us/blog/communication-success/201709/how-reduce-um-s-and-ah-s-fillers-when-speaking>, accessed 08/03/2019.
- [20] Mortensen, K.W. (2011). *The laws of charisma - How to captivate, inspire, and influence for maximum success*. New York: Amacom.
- [21] Duez, D. (2001). Acoustico-phonetic characteristics of filled pauses in spontaneous French speech: Preliminary Results. *Proc. Disfluency in Spontaneous Speech (DiSS'01), Edinburgh, UK*, 41-44.
- [22] Mobius, B. (2003). Gestalt psychology meets phonetics : An early experimental study of intrinsic F0 and intensity. *Proc. 15th ICPHS, Barcelona, Spain*, 2677–2680.
- [23] Beck, J. (2015). ExperimentMFC - Erstellung und Auswertung eines Perzeptionsexperimentes in Praat. *Kalipho* 3, 81-113.
- [24] Ménard, L., J.L. Schwartz, & L.J. Boë (2004). The role of vocal tract morphology in speech development: perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *Journal of Speech Language and Hearing Research* 47, 1059-1080.
- [25] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International* 5, 341-345.
- [26] Skarnitzl, R. & T. Boril (2015). Optimizing the Extraction of Vowel Formants. In O. Niebuhr & R. Skarnitzl (eds), *Tackling the complexity in speech* (pp. 165-182). Prague: PUP.
- [27] De Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19, 85-114.
- [28] Gósy, M, Bóna, J., Beke, A., & Horváth, V. (2014). Phonetic characteristics of filled pauses: the effects of speakers' age. *Proc. 10th International Seminar of Speech Production, Cologne, Germany*, 150–153.
- [29] Yuan J., Xiaoying X., Lai W., & Liberman M. (2016). Pauses and pause fillers in Mandarin monologue speech: The effects of sex and proficiency. *Proc. 8th International Conference of Speech Prosody, Boston, USA*, 1-5.
- [30] van der Graaf, J. (2018). *Um or Um? Language- and Speaker-specificity of Filled Pauses in Dutch and English*. BA thesis, Utrecht University, The Netherlands.
- [31] Batliner, A. et al. (1995). Filled pauses in spontaneous speech. *Proc. 13th International Congress of Phonetic Sciences, Stockholm, Sweden*, 472-475.
- [32] Wester, M., Aylett, M., Tomalin, M., & Dall, R. (2015). Artificial Personality and Disfluency. *Proc. ISCA 2015, Dresden, Germany*.